

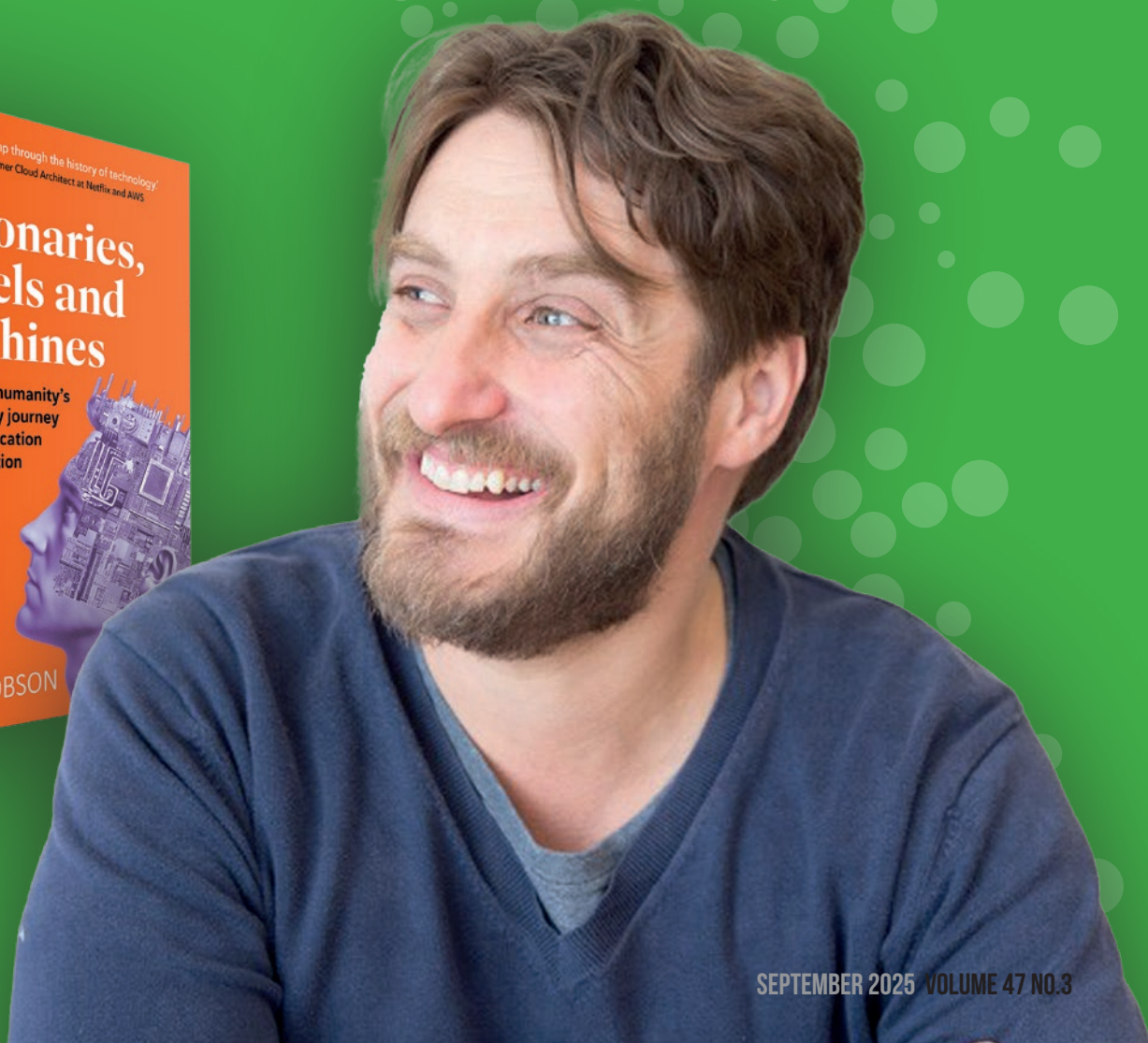
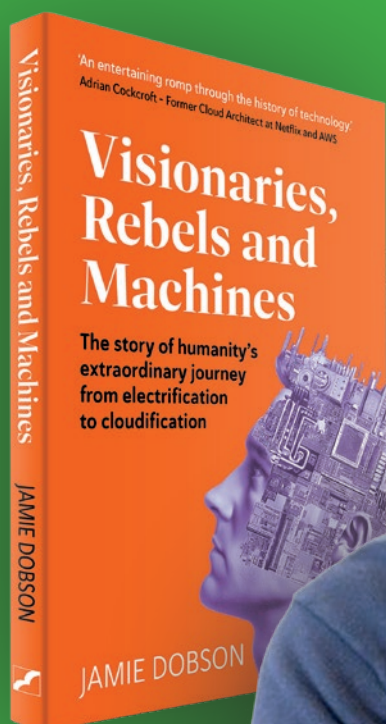
FROM THE INDUSTRY

Do We Need

TO RESTRICT AI'S HUNGER FOR DATA?

BY JAMIE DOBSON, FOUNDER OF CONTAINER SOLUTIONS AND
AUTHOR OF 'VISIONARIES, REBELS AND MACHINES'

JAMIE DOBSON, AUTHOR OF VISIONARIES, REBELS AND MACHINES, EXPLORES THE MAVERICKS AND MILESTONES THAT HAVE SHAPED THE EVOLUTION OF TECHNOLOGY. IN THIS EXCLUSIVE ARTICLE FOR BROADBAND JOURNAL, HE REFLECTS ON THE DISRUPTIVE FORCES BEHIND DIGITAL TRANSFORMATION. EXPECT BOLD INSIGHTS, SHARP STORYTELLING, AND A FRESH LOOK AT HOW WE GOT HERE.



The Industrial Revolution—and the pivotal role played by James Watt's steam engine—is often cited when discussing fundamental changes to our economic and social fabric. Since then, we've had a number of other "revolutions".

The harnessing of electricity ushered in mass production, electrified cities, and a communications revolution driven by the telegraph and later the telephone. Cars existed prior to Henry Ford's moving assembly line, but this automotive revolution made them affordable, fundamentally changing global transportation, urban planning and industry. The Space Age—beginning in 1957 with the launch of Sputnik—transformed communication and navigation, and the microprocessor, the Internet, DNA technology and mobile phones all kicked off their own revolutions.

But the Industrial Revolution is still the biggie – the most bang-for-your-buck we've had in terms of shifting human civilisation. Until now.

Artificial Intelligence is coming for James Watt's crown. And AI's revolution will be very different. Unlike previous technological revolutions that primarily transformed industries reliant on physical labour, AI's impact extends to intellectual and creative domains previously considered uniquely human.

AI'S APPETITE FOR DATA

Modern AI systems learn by digesting vast quantities of human-created content. ChatGPT, Copilot and DALL-E are sophisticated pattern-recognition systems trained on billions of examples of human creativity and knowledge.

Initially, tech companies trained these models on publicly available data, but as models grew more sophisticated, they required ever more data. Companies expanded their harvesting to include copyrighted content, paywalled articles and private repositories. And that's a problem for creators relying on compensation for their efforts, skill and talent.

Many AI companies delete their training data after models are built, making it nearly impossible to audit models for bias or track the use of copyrighted material. But even if this were not the case, currently, most jurisdictions have no specific regulations governing how companies can use publicly

available data for AI training. This regulatory vacuum has allowed AI developers to operate under a take-first-ask-later approach, creating multi-billion-dollar technology platforms using content they didn't create or license.

UNDERMINING CREATIVE INDUSTRIES

The creative sector faces a unique double threat: not only might their jobs be automated, but their existing work could be used to train the very systems replacing them. When AI systems can freely reproduce the style and substance of human creators without compensation, we're looking at a potentially destructive cycle:

1. AI systems train on human-created content without compensation
2. Economic incentives for human creation diminish
3. New content production declines in quantity or quality
4. AI systems have less novel material to learn from
5. AI outputs become increasingly derivative and homogenised

So, it's in everyone's best interests for creators to continue to produce original content by being properly compensated for their contribution to AI learning. But will this happen without regulation?

As governments worldwide grapple with these challenges, several regulatory approaches are emerging:

OPT-IN OR OPT-OUT MODELS

The simplest solution could be to create a system for opting content in or out of AI training models. In theory, this could be quick to implement with minimum complexity. Yet, given that some models are already being trained on copyrighted content (which should already be a legal "opt-out"), it might not be particularly effective.

- **Opt-Out Model (UK Proposal):** Content creators must explicitly mark their work as not available for AI training. This places the burden on creators to protect their content.

- **Opt-In Model (Advocated by Creator Organisations):** AI companies must obtain permission before using copyrighted material, similar to how music licensing works.

For businesses, an opt-out system offers fewer obstacles to AI development but creates long-term legal uncertainty. An opt-in system provides clearer legal boundaries but potentially slower access to training data.

The UK's proposed opt-out mechanism is particularly contentious. It's essentially telling creators that someone can take their property unless they explicitly post a "No Trespassing" sign – in a language that hasn't been invented yet. Critics argue this approach heavily favours large tech companies, as creators could easily lose rights to their work by simply forgetting to check a box or failing to implement technical measures they may not even understand.

Another issue is policing this and ensuring opted out data is not inadvertently or deliberately used.

DATA RIGHTS AND COMPENSATION MODELS

Similar to how music and literary rights work, content creators could receive compensation when their work is used for AI training. This could be done on an ad-hoc basis, like music streaming or through government distribution via a digital tax.

- **Collective licensing:** Creators register with collecting societies that negotiate with AI companies and distribute payments based on usage. This model exists in music with performing rights organisations such as PRS in the UK, ASCAP and BMI in the USA, GEMA in Germany or SACEM in France.
- **Data dividend:** A tax or fee on AI companies based on their data usage, with proceeds distributed to creators. This resembles public lending rights systems in countries like the UK, Canada and Australia, where authors receive payments when libraries lend their books.
- **Direct licencing:** Individual negotiations between major content producers and AI companies, with standardised terms for smaller creators.

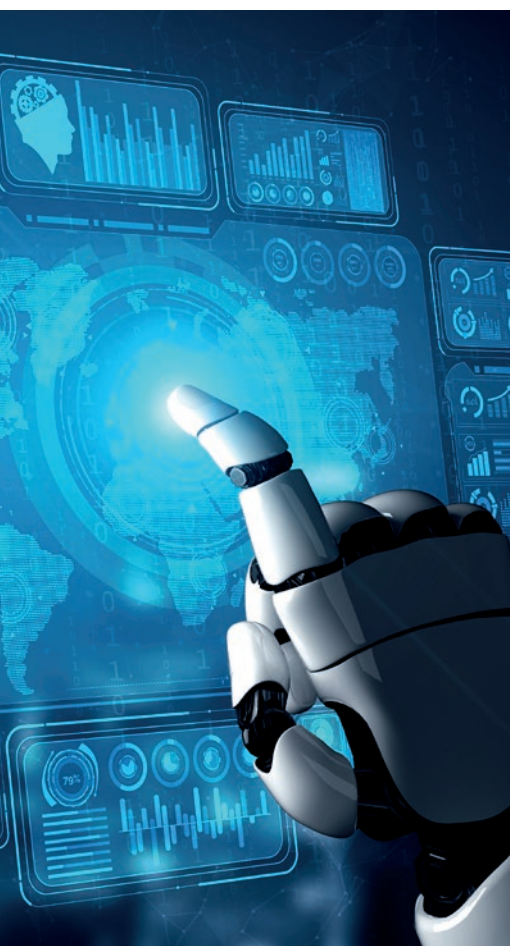
FROM THE INDUSTRY

AI AS A PUBLIC RESOURCE

Some experts advocate treating advanced AI systems like public utilities or natural monopolies. This would work similarly to electricity companies, for example, where the national grid is seen as a natural monopoly and the government implements certain standards and expectations for managing it as a public resource.

- Private companies would continue developing AI, but under enhanced regulatory oversight
- Transparency requirements would include regular audits and public reporting
- Universal access provisions would ensure broad distribution of benefits
- Price controls or licencing requirements would prevent monopolistic practices

This approach draws from how telecommunications, electricity and other essential services are regulated in many countries. It acknowledges both the innovation potential of private enterprise and the public interest in fair, accessible AI systems.



TRANSPARENCY AND TECHNICAL SAFEGUARDS

Any potential regulatory framework will require some level of transparency and technical safeguards to ensure that AI is not operating as a black box. We need to know how the algorithms are fed and on what data to ensure that creators are fairly compensated and we aren't introducing systemic biases into what will become ubiquitous technology. In other words, whatever system is chosen, it needs to be tracked and policed to ensure compliance. And this 'policing' needs to be paid for.

This will entail:

- **PROVENANCE TRACKING:** Requiring AI companies to maintain comprehensive records of their training data sources, making it possible to audit systems and verify compliance.
- **ATTRIBUTION SYSTEMS:** Building technical capabilities into AI systems that credit original creators when their work influences outputs.
- **CONTENT AUTHENTICATION:** Developing standards for watermarking or otherwise identifying AI-generated content to distinguish it from human-created work.

These technical measures would address what many legal experts consider a fundamental flaw in current AI development: the lack of transparency about what data is being used and how it influences outputs.

The publishing industry could offer a useful approach. Copyright registration systems and ISBN standards create a framework for tracking and attributing written works. Similar systems could be developed for AI training data, creating both accountability and the technical infrastructure for fair compensation.

THE INTERNATIONAL REGULATORY LANDSCAPE

Different regions are taking distinctly different approaches to AI regulation:

THE EUROPEAN UNION: COMPREHENSIVE PROTECTION

The EU's AI Act takes a risk-based approach, imposing stringent requirements for high-risk applications while allowing more flexibility for lower-risk uses.

Regarding training data, the EU leans toward an opt-in model that requires permission before using copyrighted material.

For businesses operating in Europe, this means higher compliance costs but also a clearer regulatory landscape and potentially greater consumer trust.

THE UNITED KINGDOM: INNOVATION FIRST

The UK is positioning itself as a pro-innovation hub, with proposed regulations that favour AI developers. Their controversial opt-out model places the burden on creators to explicitly mark their work as not available for AI training.

Critics argue this approach unfairly disadvantages individual creators who may lack the technical knowledge or resources to implement protection measures. It essentially tells creators that someone can use their property unless they implement specific (as yet not developed or available) technical measures – measures many creators may not understand.

THE UNITED STATES: SECTOR-SPECIFIC APPROACH

The US has thus far avoided comprehensive AI regulation, instead focusing on sector-specific guidelines and voluntary frameworks. This creates a more fragmented landscape where rules for using data may vary significantly across industries.

This approach offers flexibility but creates regulatory uncertainty for businesses, who may need to navigate different requirements across sectors and jurisdiction, which in turn increases costs and risks.

CONCLUSION

The current regulatory vacuum around AI's use of data cannot persist indefinitely. Whether through government regulation, industry self-regulation, or landmark legal cases, new frameworks for managing AI's relationship with human creativity must emerge.

The businesses that thrive won't be those extracting maximum short-term value from unregulated data harvesting, but those building sustainable models that respect and reinforce the creative ecosystem upon which AI ultimately depends.

www.container-solutions.com